Example:

I want to test a new weight loss drug. I recruit 10 randomly selected subjects and weigh them before the experiment, and then again after they have taken the weight loss drug for 30 days. The results are below. How can I test to see if the drug lowered their average weight?

Subject	1	2	3	4	5	6	7	8	9	10
Before	305	387	250	253	287	176	195	167	194	269
After	257	367	245	203	231	104	150	185	190	243

Ch. 25: Paired t-Test

** Testing the average (mean) difference μ_d

Paired Data:

* Two samples, DEPENDENT (put in L1 and L2)

Before After

* We look at the DIFFERENCES between the two sets of data
  - take 2 samples & combine to 1 sample (subtracting)
  - L3 = L2 - L1 = differences

* Do a 1-sample t-test or interval on the differences
  Test/Interval on L3

HYPOTHESES:

Ho: $\mu_d = \#$

μ_d = mean (average) difference

often 0

Ha: $\mu_d \gtrless \neq \#$

μ_d = After -- Before <-- must define what μ_d is

Conditions:

1- Paired Data — 1) data recorded on same individuals before & after

2- SRS

3- pop of differences $\geq 10n_d$

4- normal pop of differences or $n_d \geq 30$

Conditions met, t-distrib., (1 sample) paired t Test/Int.

MECHANICS:

Either a 1 sample T Interval or a 1 sample T Test

df = n -- 1

CONCLUSION:

Same conclusions from 1 sample T Interval or Test

- Reject/Fail
- Suff. evid.

* answer the question posed

** MUST say in context!

"average difference between __A__ & __B__ is ...."

μ_d

3rd sentence

DO NOT say "difference in the averages"

$\mu_A - \mu_B$

Example: A random sample of 12 sixth graders was given a memory test. They were then enrolled in a 9-month chess program. At the end of the program they were given another memory test. The researchers were interested to see if learning chess would increase memory.

Student	1	2	3	4	5	6	7	8	9	10	11	12
Pretest L_1	510	610	640	675	600	550	610	625	450	720	575	675
Posttest L_2	850	790	850	775	700	775	700	850	690	775	540	680

Diferences: $L_3 = L_2 - L_1$

Ho: $\mu_d = 0$

Ha: $\mu_d > 0$

μ_d = posttest − pretest

A - B

Hypotheses:

normal prob. plot of L_3

Conditions:

1) Paired Data

2) SRS

3) $pop_d \geq 10 n_d$

4) normal pop of diff. or $n_d \geq 30$

1) memory test taken by same students before + after program

2) stated SRS.

3) there are more than 120 sixth graders

4) normal prob. plot of diff. is roughly linear, therefore normal pop. assumed

Mechanics: cond. met, t distribution, paired t test

$$t = \frac{144.58 - 0}{\frac{109.74}{\sqrt{12}}} = 4.564 \qquad df = 11$$

$$P(t > 4.564) = 4.057 \times 10^{-4}$$

T-test
DATA
$\mu_0 = 0$
L_3
$>$

Conclusion: We reject Ho b/c p-value of $4.057 \times 10^{-4} < \alpha = 0.05$. We have suff. evid. that the average diff in memory test scores before + after the chess program is greater than 0 pts. We have evidence that the chess program worked.

Confidence Interval: (cond met -> t distrib -> paired t Int)

Mechanics: 90%

$$144.583 \pm (1.796)\left(\frac{109.74}{\sqrt{12}}\right)$$

T-Interval
DATA
L_3
C: 0.90

Conclusion: $= (87.691, 201.48)$

We are 90% conf. that the average difference in the scores on memory test before + after chess program is btw. 87.691 pts and 201.48 pts.

Example:

I want to test a new weight loss drug. I recruit 10 randomly selected subjects and weigh them before the experiment, and then again after they have taken the weight loss drug for 30 days. The results are below. The company that makes the drug claims that the patients will lose at least 30 pounds. Test this claim.

Subject	1	2	3	4	5	6	7	8	9	10
Before	305	387	250	253	287	176	195	167	194	269
After	257	367	245	203	231	104	150	185	190	243

Hypotheses: Ho: $\mu_d = 30$ $\qquad \mu_d$ = before - after

Ha: $\mu_d > 30$

Conditions:

1) Paired Data — 1) The weight measurements were taken on the same subjects before/after the drug was used

2) SRS — 2) stated random

3) pop of diff $\geq 10n_d$ — 3) all patients taking the drug ≥ 100

4) norm pop of diff or $n_d \geq 30$ — 4) Normal Probability plot looks roughly linear --> norm. pop assumed

Conditions met -> t distrib. -> 1 sample paired T test

Mechanics:

$$t = \frac{30.8 - 30}{\frac{28.09}{\sqrt{10}}} = 0.09$$

$P(t > 0.09) = 0.465$ $\qquad$ df = 9

Conclusion:

We fail to reject Ho b/c p-value of $0.465 > \alpha = 0.05$.

We have insufficient evidence that the average difference in the weight of subjects before and after the weight loss drug is more than 30 lbs, and conclude the drug did not work as claimed.

Example 1: A manufacturer wishes to compare the wearing qualities of two different types of automobile tires, A and B. For the comparison, a tire of type A and one of type B are randomly assigned and mounted on the rear wheels of each of five automobiles. The cars are then operated for a specified number of miles, and the amount of wear is recorded for each tire. Do the data present sufficient evidence to indicate a difference in the average wear for the two tire types?

a) Test the hypotheses at the 0.07 significance level.

b) Estimate the mean difference in wear by constructing a confidence interval. Use your alpha to determine what level of confidence you should use.

	1	2	3	4	5
A	10.6	9.8	12.3	9.7	8.8
B	10.2	9.4	11.8	9.1	8.3

1) Ho: $\mu_d = 0$

Ha: $\mu_d \neq 0$ μ_d = tire B - Tire A

Conditions

1) paired data — 1) the two tires were tested on the same car.

2) SRS — 2) Assumed

3) pop $\geq 10n_d$ — 3) there are more than 50 cars that will be tested

4) normal pop of diff. or $n_d \geq 30$ — 4) normal probability plot is linear --> normal population of diff.

Conditions met --> student's t-distribution --> Paired 1 sample t-test

$$t = \frac{-0.48 - 0}{0.084/\sqrt{5}} = -12.829$$

$2 * P(t < -12.829 | df = 4) = 2.128 \times 10^{-4}$

We reject Ho b/c p-value of $2.128 \times 10^{-4} < \alpha = 0.07$. We have sufficient evidence that the average difference in the wear of tires B and A is not 0 units. Therefore we have evidence of a difference between the wear on tires A and B (on average).

Conditions met --> Students t-distribution --> Paired 1 sample t interval μd = B -- A df=4

$$-0.48 \pm (2.456)\left(\frac{0.084}{\sqrt{5}}\right) = (-0.5719, -0.3881)$$

We are 93% confident that the true average difference between tire wear for brands A and B is between -0.5719 and -0.3881 units.

We are 93% confident that on average Tire A weares between 0.5719 and 0.3881 units more than Tire B.

Example 2: In response to a complaint that a particular tax assessor (A) was biased, an experiment was conducted to compare the assessor named in the compliant with another tax assessor (B) from the same office. Eight properties (1 – 8) were selected, and each was assessed by both assessors. The assessments (in thousands of dollars) are shown in the table.

a) Do the data provide sufficient evidence to indicate that assessor A tends to give higher assessments than assessor B?

b) Estimate the difference in mean assessments for the two assessors. (conditions met)

	1	2	3	4	5	6	7	8
A	76.3	88.4	80.2	94.7	68.7	82.8	76.1	79.0
B	75.1	86.8	77.3	90.6	69.1	81.0	75.3	79.1

(a) Ho: $\mu_d = 0$ μ_d = B - A

Ha: $\mu_d < 0$

Ho: $\mu_d = 0$

Ha: $\mu_d > 0$

$\mu_d = A - B$

conditions met --> t distribution --> 1 sample paired t test

$$t = \frac{-1.4875 - 0}{1.491/\sqrt{8}} = -2.8211$$

$t = 2.8211$

$P(t > 2.8211)$

$P(t < -2.8211 | df = 7) = 0.0129$

We reject Ho b/c p-value of $0.0129 < \alpha = 0.05$. We have sufficient evidence that the average difference between the assessment values of Assessor B and A is less than \$0. We have evidence that Assessor A does give higher property values.

(b) conditions met --> t-distribution --> 1 samp paired t-Int

$$-1.4875 \pm (1.895)\left(\frac{1.491}{\sqrt{8}}\right) = (-2.486, -0.4885)$$

We are 90% confident that the true average difference between the assessments of tax assessor B and A is between -2.486 and -0.4885 thousand dollars.

In other words, we are 90% confident that assessor A is on average \$2,486 to \$485 higher in his assessments compared to assessor B.